

Detection of Unlabeled Hybridized DNA and RNA Using Restriction Enzyme Digestion

Technical Field

This invention is related to the fields of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) sequencing; and genetic mapping, characterization, and diagnostics; and detection
5 and identification of organisms, viruses, and nucleic acids.

Background of the Invention

DNA and RNA – The nuclei of living cells possess chromosomes, which contain the genetic information necessary for growth, regeneration and other functions of organisms. Instructions concerning such functions are contained in the molecules of deoxyribonucleic acid (DNA). The genetic information of DNA is contained in the sequence of nucleotide bases, which are arranged on a linear polymer of deoxyribose phosphates. The four bases are thymine (T), adenine (A), cytosine (C), and guanine (G). Two strands of DNA form a double helix joined by base pairing of Ts to As and Cs to Gs. Accordingly, the base sequence along one strand
15 determines the order of bases along the complementary strand. Cells use DNA as a template for RNA synthesis. During RNA synthesis, a strand of RNA complementary to the template DNA is formed. RNA/DNA hybrids obey the same base pairing rules as DNA/DNA hybrids except that RNA contains uracil (U) instead of T. Both RNA and DNA, therefore, can be used to derive genetic information.

DNA and RNA Sequencing – Sequencing is the determination of the sequence of bases in nucleic acid strands. Polymerase chain reaction (PCR) is a process in which segments of sample
20

DNA that are of interest are synthesized repeatedly until an amount is produced which is suitable for further experimentation. For the purpose of sequencing, PCR can contain fluorescently labeled, dideoxynucleotide triphosphates that stop further elongation of strands. Each of the four kinds of dideoxynucleotide triphosphates (A, T, C, or G) is labeled with a different fluorescent label. The products of this reaction are of various lengths and each strand has only one label that represents the last base added. When separated by size, the base sequence of the template DNA can be determined. This is the Sanger method of DNA sequencing and is, at present, the most commonly used method for sequencing long DNA sequences. For RNA, PCR includes reverse transcriptase that makes a complementary DNA from the RNA template. The DNA products are then sequenced by the Sanger method. The Sanger method requires a high quality gel and electrophoresis of the amplified DNA. It is a time consuming process and not well-suited for parallel sequencing of several sequences.

Sequencing by Hybridization – Sequencing by hybridization (SBH) uses the specificity of DNA/DNA binding rules and melting temperature to sequence labeled sample DNA. Hybridization is performed at elevated temperatures close to the melting point of the fully complementary hybrid so that non-complementary DNA cannot hybridize to the probe. Also, imperfect hybridization can be followed by washes approaching the melting temperature to ensure that non-complementary sample DNA has been removed from the probe site. Typically, probes are immobilized on a solid surface. Detection of labeled DNA at the location of a probe indicates the presence of complementary sequence in the sample DNA. The solid surface containing the probe oligonucleotides is sometimes referred to as an SBH chip, a genosensor chip, a hybridization surface, or a dot blot. This version of the SBH process is referred to as

Format II SBH. E. Southern disclosed Format II SBH in International Application No. PCT/GB89/00460.

Format I SBH is an alternative method where sample DNA is attached to a solid surface, such as a nylon membrane, and hybridized with labeled probes. R. Drmanac and R.

5 Czerkvenjakov disclosed Format I SBH in US Patent No. 5,202,231.

Genetic diagnostic information can also be obtained using techniques similar to those used for obtaining sequencing information. In diagnostic applications, a known sequence that indicates the presence of a particular gene or organism is immobilized to the surface and labeled DNA from the sample is allowed to hybridize. After washing to eliminate nonspecific binding, detection of the label indicates that gene or organism is present in the sample.

While the above sequencing by hybridization techniques are capable of producing reliable results, if properly applied, certain disadvantages are inherent. For example, the target DNA must be sorted and immobilized (Format I) or labeled (Format II). These procedures for sequencing are laborious and require a large DNA or RNA sample. If these limitations can be overcome, the parallel nature of the genosensor chip makes it ideal for several forms of diagnostics, forensics, and other applications. In particular, if amplification, labeling, and segregation of the target DNA before hybridization were not necessary, the genosensor chip could be used more readily as a routine diagnostic test. DNA arrays utilize parallel hybridization of target DNA to multiple known-sequence probes in a compact array. This allows detection of multiple sequences using SBH with only one small DNA sample [1-8]. Since the arrays are so compact (e.g., 100s to 1000s on a silicon chip), the reliability of diagnostic tests can be improved by using multiple probes.

Probe Technology - Molecular beacons (described in US Patents 5,118,801 and 5,312,728) are single-stranded DNA (ss-DNA) probes that form a stem-and-loop structure. The loop sequence is complementary to a sequence of interest. Binding of complementary DNA to the loop sequence opens the stem-and-loop structure. If no complementary DNA is present, the stem-and-loop structure remains intact. Molecular beacons do not require labeling of the sample DNA because the beacons have a single fluorophore attached to one end and a quencher to the other end. When in a stem-and-loop conformation, the ends are close together and light emission is quenched. When hybridization breaks the stem structure, the fluorophore and quencher are separated and light is emitted [9]. Although the stem is only 6 or 7 base pairs in length, it has a melting temperature well above 55°C due to its intramolecular nature. Molecular beacons are routinely used to monitor PCR reactions in real time.

Molecular beacons can be used to differentiate sequences that differ by only a single base such as alleles for drug-resistant tuberculosis, human coagulation factor V, human estrogen receptor, and human methylenetetrahydrofolate reductase [10-13]. This specificity is not available with most other probe types. In addition to DNA studies, molecular beacons have been used to detect RNA sequences in living cells [14-16]. It has been demonstrated that molecular beacons still function when immobilized to surfaces [17-18].

It is necessary that molecular beacons form the desired stem-and-loop structure. One useful tool is a DNA folding program available on the internet (<http://bioinfo.math.rpi.edu/~zukerm/>) that estimates the free energy of formation of the stem hybrid and predicts the melting temperature. In general, the loop areas should be approximately 21 bases long with a GC content of 25-75 percent. The loop length can be adjusted to

accommodate various GC percentages. The stem structures should be 6 or 7 base pairs long and contain mainly Cs and Gs.

Although molecular beacons have many good qualities, they are limited for use in applications other than PCR diagnostics. The use of a single fluorophore label requires the amplification of the sample DNA. Because many other compounds fluoresce and produce a high background, a large amount of fluorophore must be present to be measured accurately.

Label Technology - Bioluminescent enzymes, such as luciferase, alkaline phosphatase, horseradish peroxidase, β -galactosidase, and aequorin, have recently been used to detect attomolar concentrations of molecules and have been shown to be 30-60 fold more sensitive than radioactive labels [19]. They produce much less background since no excitation light is required. In these particular cases, light is produced continuously as long as activator is available, therefore, photons can be counted over a period of time. Label technology is a rapidly advancing field, and, for this reason, a type of probe that is easily adaptable to new labels or multiple labels would be highly beneficial.

Restriction enzymes – Restriction enzymes (or restriction endonucleases) are produced in bacteria, presumably to degrade foreign DNA. Methylation differences between the bacterium's genomic DNA and the foreign DNA protect the genomic DNA from cleavage [20].

Restriction enzymes bind at recognition sequences. Recognition sequences are typically 4 to 6 bases long, but may be longer. Restriction enzymes cleave double-stranded DNA (dsDNA) at a restriction site, which may or may not be located within the recognition sequence. At each restriction site, one phosphodiester bond from each of the strands in the dsDNA is hydrolyzed to form hydroxyl and phosphate groups. The cleaved sites, one on each DNA strand, may be

opposite each other forming two blunt-ended dsDNA fragments, or may occur at different locations resulting in fragments with protruding unpaired bases called sticky ends [21].

Currently, almost 3000 restriction enzymes with over 200 different recognition sequences are known. The field of molecular cloning relies heavily on the use of restriction enzymes to create new chimeric DNA sequences from different sources [22]. This demand is met by several companies (New England Biolabs, Fermentas, Roche, and Stratagene, for example) that offer restriction enzymes and study the activities of the enzymes. Restriction enzymes are usually supplied with a concentrated buffer solution that optimizes activity. Typical buffer conditions are pHs from 7.0 to 7.9, magnesium salts from 10 to 100 mM, bovine serum albumin from 0 to 100 µg/ml, and temperatures from 30 to 65°C. One unit of restriction enzyme is defined as the amount necessary to digest 1 µg of DNA in one hour under optimal conditions.

The most commonly used restriction enzymes are highly efficient (nearly 100% of restriction sites cut). DNA fragment length can affect restriction enzyme activity. Frequently, restriction enzymes can cut short pieces of dsDNA efficiently, but some require longer fragments. Suppliers (New England Biolabs, for example) recommend having six base pairs on each side of recognition sites to ensure efficient cleavage when length preference is unknown. Some enzymes, such as *Eco* RI, are known to cut efficiently with only one base on each side of the recognition site.

Summary of Invention

The present invention utilizes a nucleic acid probe for detecting specific target nucleic acid sequences. The probe is an oligonucleotide containing a covalent surface-coupling group attached to one end and a label attached to the other end, wherein the oligonucleotide contains

two complementary regions which hybridize to each other in the absence of target nucleic acid forming a loop region which contains a sequence complementary to the target sequence being probed, and forming a stem region which contains a restriction enzyme site which is not present in the hybridized loop.

5 The present invention further utilizes a method of determining the presence of one or more specific target nucleic acid sequences using such a probe.

Brief Description of the Drawings

10 **Figure 1.** The structure of a probe with a removable label (PRL) attached to a glass surface. G, C, A, and T represent the bases of nucleotides in an ssDNA molecule. The looped area contains the target-hybridization sequence. Stem-forming sequences, located on each side of the target-hybridization sequence, hydrogen bond to form the stem. This stem can be cut by the restriction enzyme *Pal* I.

15 **Figure 2.** A summary of the processes in a preferred embodiment.

Detailed Description of the Invention

20 The invention is a probe with a removable label (PRL) that can be used to detect hybridization without labeling target DNA. The PRL structure differs from that of a molecular beacon in three ways: 1) the stem structure contains a restriction enzyme recognition site, 2) different label types, including fluorescent labels, can be used and no quencher is necessary, 3) a moiety is present to attach the PRL to surfaces to form arrays (Figure 1). Like molecular beacons, the stem structure in the PRL must open upon hybridization of the loop sequence to

complementary target DNA or RNA. The stem structures may be approximately 4-9 nucleotides in length. For each PRL array, a restriction enzyme is used that has a restriction site in each stem sequence but does not have a restriction site or a recognition site in any of the hybridized loop sequences. A list of several potential restriction enzymes and their recognition sequences is given

5 in Table 1. Identification of additional suitable restriction enzymes can be readily determined based on the present disclosure of the invention and generally available knowledge in the field of molecular biology. For example, the PRL shown in Figure 1 has both the *Pal* I recognition site and restriction site located in the stem. The stem sequence GGCCAG, written 5' to 3', can be cut by *Pal* I forming the fragments GG and CCAG. After this fragmentation, covalent bonds no

10 longer connect the label to the surface. The six weaker hydrogen bonds between the two CG base pairs are not strong enough to hold the label to the bound portion of the PRL and the label can be washed away. Other stem sequences, used with other restriction enzymes in Table 1, may have more than six hydrogen bonds remaining. Washes at elevated temperatures can be used to break these hydrogen bonds without damaging the covalent bonds. The conditions used for washing

15 away the labeled nucleotide fragment subsequent to restriction enzyme digestion may be routinely determined by one skilled in the art based on the restriction site and restriction enzyme used. Such information is typically available from the company source of the restriction enzyme.

The loop portion of the PRL should be about 16 to 25 nucleotides in length. In order to function properly, the hybrid formed upon hybridization of a fully complementary DNA to the

20 loop sequence must be more energetically favorable than the stem hybrid, however, any loop hybrid containing mismatches must be energetically less favorable than the stem hybrid. In order to achieve this, the length or composition of the loop and /or stem sequences is varied. Commonly, a loop sequence with a high AT content requires more nucleotides in the loop

sequence or a weaker stem sequence. A loop sequence with a higher GC content may need a reduced length or a more stable stem sequence.

Table 1. A list of restriction enzymes and their recognition sequences. Abbreviations are as follows: ^, indicates the cleavage site; _, indicates the second cleavage site in "sticky end" cleavages; N, any of the four bases; W, base A or T; M, base A or C; K, base G or T; Y, base C or T; R, base G or A; S base G or C; D base A or G or T; H base A or C or T. Table adapted from the Restriction Enzyme Database (<http://rebase.neb.com>), Dr. Richard J. Roberts and Dana Macelis, authors.

<u>Enzymes</u>	Recognition Sequence	Other Enzymes Capable of Cutting These Sequences
<i>AluI</i>	AG^CT	<i>MltI</i>
<i>BalI</i>	TGG^CCA	<i>MlsI</i> , <i>Mlu31I</i> , <i>MluNI</i> , <i>MscI</i>
<i>BsaAI</i>	YAC^GTR	<i>BstBAI</i> , <i>MspYI</i> , <i>PsuAI</i>
<i>BsrBI</i>	CCG^CTC	<i>AccBSI</i> , <i>BstD102I</i> , <i>Bst31NI</i> , <i>MbiI</i>
<i>BtrI</i>	CAC^GTC	-
<i>Cac8I</i>	GCN^NGC	-
<i>CviJI</i>	RG^CY	<i>CviTI</i>
<i>CviRI</i>	TG^CA	<i>HpyCH4V</i>
<i>Eco47III</i>	AGC^GCT	<i>AfeI</i> , <i>AitI</i> , <i>Aor51HI</i> , <i>FunI</i>
<i>Eco78I</i>	GGC^GCC	<i>EgeI</i> , <i>EheI</i> , <i>SfoI</i>
<i>EcoICRI</i>	GAG^CTC	<i>Ecl136II</i> , <i>Eco53kI</i> , <i>MxaI</i>
<i>FnuDII</i>	CG^CG	<i>AccII</i> , <i>BceBI</i> , <i>BepI</i> , <i>Bpu95I</i> , <i>Bsh1236I</i> , <i>Bsp50I</i> , <i>Bsp123I</i> , <i>BstFNI</i> , <i>BstUI</i> , <i>Bsu1532I</i> , <i>BtkI</i> , <i>Csp68KVI</i> , <i>CspKVI</i> , <i>FauBII</i> , <i>MvnI</i> , <i>ThaI</i>
<i>HaeI</i>	WGG^CCW	-
<i>HaeIII</i>	GG^CC	<i>BecAII</i> , <i>Bim19II</i> , <i>Bme361I</i> , <i>BseQI</i> , <i>BshI</i> , <i>BshFI</i> , <i>Bsp211I</i> , <i>BspBRI</i> , <i>BspKI</i> , <i>BspRI</i> , <i>BsuRI</i> , <i>BteI</i> , <i>CltI</i> , <i>DsaII</i> , <i>FnuDI</i> , <i>MchAII</i> , <i>MfoAI</i> , <i>NgoPII</i> , <i>NspLKI</i> , <i>Pall</i> , <i>Pde133I</i> , <i>PfIKI</i> , <i>PlaI</i> , <i>SbvI</i> , <i>SfaI</i> , <i>SuaI</i>
<i>HindII</i>	GTY^RAC	<i>HinJCI</i> , <i>HincII</i>
<i>Hpy8I</i>	GTN^NAC	<i>HpyBII</i>
<i>LpnI</i>	RGC^GCT	<i>Bme142I</i>
<i>MstI</i>	TGC^GCA	<i>Acc16I</i> , <i>AosI</i> , <i>AviII</i> , <i>FdiII</i> , <i>FspI</i> , <i>NsbI</i> , <i>PamI</i>
<i>NaeI</i>	GCC^GGC	<i>CcoI</i> , <i>PdiI</i> , <i>SauBMKI</i> , <i>SauHPI</i> , <i>SauLPI</i> , <i>SauNI</i> , <i>SauSI</i> , <i>Slu177I</i>
<i>NlaIV</i>	GGN^NCC	<i>AspNI</i> , <i>BscBI</i> , <i>BspLI</i> , <i>PspN4I</i>
<i>NruI</i>	TCG^CGA	<i>Bsp68I</i> , <i>MluB2I</i> , <i>Sbo13I</i> , <i>SpoI</i>
<i>NspBII</i>	CMG^CKG	<i>MspAII</i>
<i>PmaCI</i>	CAC^GTG	<i>BbrPI</i> , <i>BcoAI</i> , <i>Eco72I</i> , <i>PmiI</i>

<u>Enzymes</u>	<u>Recognition Sequence</u>	<u>Other Enzymes Capable of Cutting These Sequences</u>
<i>PvuII</i>	CAG [^] CTG	<i>BavI</i> , <i>BavAI</i> , <i>BavBI</i> , <i>Bsp153AI</i> , <i>BspM39I</i> , <i>BspO4I</i> , <i>Cfr6I</i> , <i>DmaI</i> , <i>EciI</i> , <i>NmeRI</i> , <i>Pae17kI</i> , <i>Pvu84II</i>
<i>RsaI</i>	GT [^] AC	<i>AfaI</i> , <i>HpyBI</i> , <i>PlaAII</i>
<i>SciI</i>	CTC [^] GAG	-
<i>SmaI</i>	CCC [^] GGG	<i>CfrJ4I</i> , <i>PaeBI</i> , <i>PspALI</i>
<i>SrfI</i>	GCCC [^] GGGC	-
<i>SspI</i>	AAT [^] ATT	-
<i>StuI</i>	AGG [^] CCT	<i>AatI</i> , <i>AspMI</i> , <i>Eco147I</i> , <i>GdiI</i> , <i>Pme55I</i> , <i>SarI</i> , <i>Sru30DI</i> , <i>SseBI</i> , <i>SteI</i>
Type II restriction enzymes cleaving outside their recognition sequences:		
<i>BsrI</i>	ACTG_GN [^]	<i>BseII</i> , <i>BseNI</i> , <i>BsrSI</i> , <i>Bst11I</i> , <i>TspII</i>
<i>BstF5I</i>	GGATG_NN [^]	<i>BseGI</i>
<i>BtsI</i>	GCA GTG_NN [^]	-
Type II restriction enzymes producing 5' overhangs:		
<i>AccI</i>	GT [^] MK_AC	<i>FblI</i> , <i>XmiI</i>
<i>AciI</i>	C [^] CG_C	-
<i>AcyI</i>	GR [^] CG_YC	<i>AhaII</i> , <i>AosII</i> , <i>AstWI</i> , <i>AsuIII</i> , <i>BbiII</i> , <i>BsaHI</i> , <i>BstACI</i> , <i>HgiI</i> , <i>HgiDI</i> , <i>HgiGI</i> , <i>HgiHII</i> , <i>Hin1I</i> , <i>Hsp92I</i> , <i>Msp17I</i> , <i>PamII</i>
<i>AflIII</i>	A [^] CRYG_T	-
<i>AgeI</i>	A [^] CCGG_T	<i>AsiAI</i> , <i>BshTI</i> , <i>CspAI</i> , <i>PinAI</i>
<i>ApaLI</i>	G [^] TGCA_C	<i>Alw44I</i> , <i>SnoI</i> , <i>VneI</i>
<i>AscI</i>	GG [^] CGCG_CC	-
<i>Asp718I</i>	G [^] GTAC_C	<i>Acc65I</i> , <i>AhaB8I</i> , <i>SthI</i>
<i>AsuI</i>	G [^] GNC_C	<i>AspS9I</i> , <i>AvcI</i> , <i>Bac36I</i> , <i>Bal228I</i> , <i>BavAII</i> , <i>BavBII</i> , <i>Bce22I</i> , <i>BshKI</i> , <i>BsiZI</i> , <i>Bsp1894I</i> , <i>BspBII</i> , <i>BspF4I</i> , <i>Bsu54I</i> , <i>CcuI</i> , <i>Cfr13I</i> , <i>MaeK81II</i> , <i>NspIV</i> , <i>Nsp7121I</i> , <i>Pde12I</i> , <i>PspPI</i> , <i>Sau96I</i>
<i>AsuII</i>	TT [^] CG_AA	<i>Acpl</i> , <i>Asp10HI</i> , <i>BimI</i> , <i>Bim19I</i> , <i>Bpu14I</i> , <i>BsiCI</i> , <i>Bsp119I</i> , <i>BspLAI</i> , <i>BspT104I</i> , <i>BstBI</i> , <i>CbiI</i> , <i>Csp45I</i> , <i>Csp68KII</i> , <i>FspII</i> , <i>LspI</i> , <i>MlaI</i> , <i>NspV</i> , <i>PlalI</i> , <i>PpaAI</i> , <i>SfuI</i> , <i>Ssp1I</i> , <i>SspRFI</i> , <i>SviI</i>
<i>AvaI</i>	C [^] YCGR_G	<i>Ama87I</i> , <i>AquI</i> , <i>BcoI</i> , <i>Bse15I</i> , <i>BsoBI</i> , <i>BstSI</i> , <i>Eco88I</i> , <i>Eco27kI</i> , <i>NspIII</i> , <i>NspSAI</i> , <i>PlaAI</i> , <i>PunAI</i>
<i>AvaII</i>	G [^] GWC_C	<i>AflI</i> , <i>Asp745I</i> , <i>BamNxi</i> , <i>BcuAI</i> , <i>Bme18I</i> , <i>Bme216I</i> , <i>BsrAI</i> , <i>BthAI</i> , <i>CauI</i> , <i>Csp68KI</i> , <i>DsaIV</i> , <i>EagMI</i> , <i>Eco47I</i> , <i>ErpI</i> , <i>FdiI</i> , <i>FspMSI</i> , <i>FssI</i> , <i>HgiBI</i> , <i>HgiCII</i> , <i>HgiEI</i> , <i>HgiHIII</i> , <i>HgiJI</i> , <i>Kzo49I</i> , <i>SinI</i> , <i>SmuEI</i> , <i>VpaK11BI</i>
<i>AvrII</i>	C [^] CTAG_G	<i>AvrBII</i> , <i>BlnI</i> , <i>BspA2I</i> , <i>XmaII</i>
<i>BamHI</i>	G [^] GATC_C	<i>AccEBI</i> , <i>Alil</i> , <i>ApaCI</i> , <i>AsiI</i> , <i>Bce751I</i> , <i>BnaI</i> , <i>Bsp98I</i> , <i>Bsp4009I</i> , <i>BspAAIII</i> , <i>BstI</i> , <i>CelI</i> , <i>Mlu23I</i> , <i>Nsp29132II</i> , <i>NspSAIV</i> , <i>OkrAI</i>

Enzymes	Recognition Sequence	Other Enzymes Capable of Cutting These Sequences
		<i>Pfl8I, RspLKII, SotI, SurI</i>
<i>BbvCI</i>	CC [^] TCA_GC	<i>AbeI</i>
<i>BeII</i>	W [^] CCGG_W	<i>Bca77I, BsaWI</i>
<i>Bpu10I</i>	CC [^] TNA_GC	<i>BpuDI</i>
<i>BsePI</i>	G [^] CGCG_C	<i>BssHII, BstBZ153I, PaeI</i>
<i>BsiI</i>	C [^] ACGA_G	<i>BssSI, Bst2BI</i>
<i>Bsp120I</i>	G [^] GGCC_C	<i>PspOMI</i>
<i>BspMII</i>	T [^] CCGG_A	<i>AccIII, Aor13HI, BbvAIII, BseAI, BsiMI, Bsp13I, BspEI, Bsu23I, CauB3I, Kpn2I, MroI, PinBII, PtaI</i>
<i>BstNI</i>	CC [^] W_GG	<i>AeuI, AorI, ApaORI, ApyI, BciBII, Bse16I, Bse17I, Bse24I, BseBI, BshGI, BsiLI, BspNI, Bst1I, Bst2I, Bst38I, Bst100I, BstM6I, BstOI, Bst2UI, BthDI, BthEI, CbrI, CthII, Fsp1604I, MvaI, SniI, SsiI, Sth117I, TaqXI, ZaiI</i>
<i>BstEII</i>	G [^] GTNAC_C	<i>AcrII, AspAI, Bse64I, BseT9I, BseT10I, BsiKI, BstPI, BstT9I, BstT10I, EcaI, Eci125I, Eco91I, EcoO65I, EcoO128I, NspSAII, PspEI</i>
<i>CauII</i>	CC [^] S_GG	<i>AhaI, AseII, AsuC2I, BcnI, HgiS22I, Mgl1448II, NciI</i>
<i>CfrI</i>	Y [^] GGCC_R	<i>Bfi89I, EaeI, EcoHK31I</i>
<i>Cfr10I</i>	R [^] CCGG_Y	<i>Bco118I, Bse118I, Bse634I, BsrFI, BssAI, BstB7SI</i>
<i>CviAII</i>	C [^] AT_G	-
<i>CviQI</i>	G [^] TA_C	<i>Csp6I, CviRII</i>
<i>DdeI</i>	C [^] TNA_G	<i>BstDEI, BstJZ30II</i>
<i>DraII</i>	RG [^] GNC_CY	<i>EcoO109I</i>
<i>DsaI</i>	C [^] CRYG_G	<i>BstDSI, BtgI, TspBI</i>
<i>EcoHI</i>	[^] CCSGG_	<i>Eco1831I, Kpn49kII</i>
<i>EcoRII</i>	[^] CCWGG_	<i>PspGI, SliI, SspAI</i>
<i>Eco56I</i>	G [^] CCGG_C	<i>MroNI, NgoAIV, NgoMIV, SceIII, SrfI</i>
<i>EspI</i>	GC [^] TNA_GC	<i>BlpI, Bpu1102I, Bsp1720I, CelII</i>
<i>Fnu4HI</i>	GC [^] N_GC	<i>BsoFI, Bsp6I, FbrI, Fsp4HI, ItaI, Sati, Uur960I</i>
<i>GdiII</i>	C [^] GGCC_R	-
<i>HgiCI</i>	G [^] GYRC_C	<i>AccBII, BanI, BbvBI, BshNI, BspT107I, Eco64I, HgiHI, MspB4I, PfaAI</i>
<i>HinI</i>	G [^] ANT_C	<i>CviBI, FnuAI, HhaII, SscLII</i>
<i>HinPII</i>	G [^] CG_C	<i>Hin6I, HsoI, HspAI, SciNI</i>
<i>HpaII</i>	C [^] CG_G	<i>Bco27I, BsiSI, Bst40I, BsuFI, Cbol, HapII, Hin2I, MnoI,</i>

<u>Enzymes</u>	<u>Recognition Sequence</u>	<u>Other Enzymes Capable of Cutting These Sequences</u>
		<i>MspI, Pde137I, Sth134I</i>
<i>Hpy178III</i>	TC [^] NN_GA	<i>Hpy188III</i>
<i>KasI</i>	G [^] GCGC_C	-
<i>MaeI</i>	C [^] TA_G	<i>BfaI, FgoI, MthZI, RmaI, XspI</i>
<i>MaeII</i>	A [^] CG_T	<i>HpyCH4IV</i>
<i>MaeIII</i>	[^] GTNAC_	-
<i>MboI</i>	[^] GATC_	<i>AspMDI, Bce243I, Bfi57I, Bme12I, BscFI, Bsp67I, Bsp105I, Bsp143I, Bsp2095I, BspAI, BspFI BspJI, BstENII, BtkII, CacI, CcyI, CpfI, CviAI, DpnII, FnuCI, FnuEI, HaeI, Kzo9I, LlaAI, MgoI, MkrAI, NdeII, NlaII, NmeCI, NphI, RalF40I, Sau3AI, SauMI, Sth368I</i>
<i>MluI</i>	A [^] GCGG_T	<i>Bbi24I</i>
<i>NarI</i>	GG [^] CG_CC	<i>MchI, Mly113I, NdaI, NunII, SseAI</i>
<i>NcoI</i>	C [^] CATG_G	<i>Bsp19I</i>
<i>NheI</i>	G [^] CTAG_C	<i>AsuNHI, PstNHI</i>
<i>NotI</i>	GC [^] GGCC_GC	<i>CciNI, CspBI, MchAI</i>
<i>PpuMI</i>	RG [^] GWC_CY	<i>Pfl27I, PpuXI, Psp5II, PspPPI</i>
<i>RsrII</i>	CG [^] GWC_CG	<i>CpoI, CspI, Rsr2I</i>
<i>SalI</i>	G [^] TCGA_C	<i>BspMKI, HgiCIII, HgiDII, NopI, RflFI, RtrI, Rtr63I, XciI</i>
<i>SanDI</i>	GG [^] GWC_CC	<i>Sse1825I</i>
<i>SauI</i>	CC [^] TNA_GG	<i>AocI, AxyI, BliHki, Bse2II, BspR7I, Bsu36I, CvnI, Eco8II, Lmu60I, MstII, OxaNI, SshAI</i>
<i>ScrFI</i>	CC [^] N_GG	<i>Bme1390I, Msp67I, MspR9I</i>
<i>SecI</i>	C [^] CNNG_G	<i>BsaJI, BseDI, BssECI</i>
<i>SeII</i>	[^] GCGG_	-
<i>SexAI</i>	A [^] CCWGG_T	<i>MabI</i>
<i>SfeI</i>	C [^] TRYA_G	<i>BdiSI, BfmI, BstSFI, LlaBI, SfeI</i>
<i>SgrAI</i>	CR [^] CCGG_YG	-
<i>SimI</i>	GG [^] GTC_	-
<i>SmII</i>	C [^] TYRA_G	-
<i>SplI</i>	C [^] GTAC_G	<i>BpuB5I, BsiWI, BvuBI, MaeK81I, Pfl23II, PpuAI, PspLI, SunI</i>
<i>Sse232I</i>	CG [^] CCGG_CG	-
<i>SsoII</i>	[^] CCNGG_	<i>BssKI, BstMZ611I, DsaV, Ecl18kI, Eco13kI, Eco21kI,</i>

<u>Enzymes</u>	<u>Recognition Sequence</u>	<u>Other Enzymes Capable of Cutting These Sequences</u>
		<i>Eco137kI, Kpn2kI, StyD4I</i>
<i>StyI</i>	C [^] CWWG_G	<i>BsmSI, BssT1I, CfrBI, Eco130I, EcoT14I, ErhI, ErhB9II</i>
<i>TaqI</i>	T [^] CG_A	<i>PpaAII, Tsp32I, Tsp32II, TthHB8I</i>
<i>TseI</i>	G [^] CWG_C	<i>AceI, Taq52I</i>
<i>Tsp45I</i>	[^] GTSAC_	<i>Hpy51I, NmuCI</i>
<i>UnbI</i>	[^] GGNCC_	-
<i>VpaK11AI</i>	[^] GGWCC_	-
<i>XhoI</i>	C [^] TCGA_G	<i>AbrI, BluI, BspAAI, BssHI, BstVI, CcrI, MavI, MlaAI, PaeR7I, PanI, Sau3239I, SauLP1I, Sbi68I, Sfr274I, SlaI, Sol10179I, StrI, TliI, XpaI</i>
<i>XhoII</i>	R [^] GATC_Y	<i>BloHI, BstX2I, BstYI, DsaIII, MffII, PstI, Tru201I</i>
<i>XmaI</i>	C [^] CCGG_G	<i>AhyI, Cfr9I, EaeAI, EclRI, Pac25I, PspAI, XcyI, XmaCI</i>
<i>XmaIII</i>	C [^] GGCC_G	<i>AaaI, BseX3I, BstZI, EagI, EclXI, Eco52I, SenPT16I</i>
Type II restriction enzymes producing 3' overhangs:		
<i>AatII</i>	G_ACGT [^] C	<i>Ssp5230I</i>
<i>AceII</i>	G_CTAG [^] C	-
<i>ApaI</i>	G_GGCC [^] C	<i>PpeI</i>
<i>BbeI</i>	G_GCGC [^] C	-
<i>BseSI</i>	G_KGCM [^] C	-
<i>BspKT6I</i>	G_AT [^] C	-
<i>BsrI</i>	ACTG_GN [^]	<i>BseII, BseNI, BsrSI, Bst11I, Tsp1I</i>
<i>BstF5I</i>	GGATG_NN [^]	<i>BseGI</i>
<i>BstXI</i>	CCAN_NNNN [^] NTGG	<i>BstHZ55I</i>
<i>ChaI</i>	_GATC [^]	-
<i>FmuI</i>	G_GNC [^] C	-
<i>FseI</i>	GG_CCGG [^] CC	-
<i>HaeII</i>	R_GCGC [^] Y	<i>AccB2I, Bsp143II, BstH2I</i>
<i>HgiAI</i>	G_WGCW [^] C	<i>Alw21I, AspHI, Bbv12I, Bsh45I, BsiHKAI, MspV281I</i>
<i>HgiIII</i>	G_RGCY [^] C	<i>BanII, BpuI, Bsp519I, Bsu1854I, BvuI, Eco24I, Eco75KI, EcoT38I, FriOI, KoxII, PaeHI, SacNI</i>
<i>HhaI</i>	G_CG [^] C	<i>AspLEI, BspLAI, BstHHI, CfoI, FnuDIII</i>
<i>Hpy99I</i>	_CGWCG [^]	-

Enzymes	Recognition Sequence	Other Enzymes Capable of Cutting These Sequences
<i>Hpy</i> 188I	TC_N^GA	-
<i>Hpy</i> CH4I	C_AT^G	-
<i>Kpn</i> I	G_GTAC^C	-
<i>Mcr</i> I	CG_RY^CG	<i>Bsa</i> OI, <i>Bsh</i> 1285I, <i>Bsi</i> EI, <i>Bst</i> MCI
<i>Nla</i> III	_CATG^	<i>Hin</i> II, <i>Hsp</i> 92II
<i>Nli</i> 3877I	C_YCGR^G	-
<i>Nsp</i> I	R_CATG^Y	<i>Bst</i> NSI, <i>Nsp</i> HI, <i>Pau</i> AI, <i>Pun</i> AII, <i>Xce</i> I
<i>Psp</i> 03I	G_GWC^C	-
<i>Pss</i> I	RG_GNC^CY	<i>Kaz</i> 48kI
<i>Pst</i> I	C_TGCA^G	<i>Ajo</i> I, <i>Ali</i> AJI, <i>Api</i> I, <i>Asp</i> 713I, <i>Bsp</i> 63I, <i>Bsp</i> BI, <i>Bsu</i> BI, <i>Cfi</i> II, <i>Cfr</i> A4I, <i>Cfu</i> II, <i>Cst</i> I, <i>Ecl</i> 37kI, <i>Ecl</i> 2zI, <i>Hal</i> II, <i>Mha</i> AI, <i>Pae</i> PI, <i>Pfl</i> 21I, <i>Psp</i> 23I, <i>Sal</i> PI, <i>Sfi</i> II, <i>Yen</i> I
<i>Pvu</i> I	CG_AT^CG	<i>Afa</i> 22MI, <i>Afa</i> 16RI, <i>Bsp</i> CI, <i>Eag</i> BI, <i>Erh</i> B9I, <i>Mvr</i> I, <i>Nb</i> II, <i>Ple</i> 19I, <i>Psu</i> 161I, <i>Rsh</i> I, <i>Xor</i> II
<i>Sac</i> I	G_AGCT^C	<i>Psp</i> 124BI, <i>Sst</i> I
<i>Sac</i> II	CC_GC^GG	<i>Cfr</i> 42I, <i>Csc</i> I, <i>Eae</i> 46I, <i>Eco</i> 29kI, <i>Ga</i> II, <i>Gce</i> I, <i>Gce</i> GLI, <i>Kpn</i> 378I, <i>Ksp</i> I, <i>Ngo</i> AIII, <i>Ngo</i> PIII, <i>Pae</i> AI, <i>Pae</i> QI, <i>Pae</i> 5kI, <i>Pae</i> 14kI, <i>Sch</i> ZI, <i>Sen</i> PT14bI, <i>Sex</i> BI, <i>Sex</i> CI, <i>Sfr</i> 303I, <i>Sgr</i> BI, <i>Spu</i> I, <i>Sst</i> II
<i>Sdu</i> I	G_DGCH^C	<i>Aoc</i> II, <i>Bmy</i> I, <i>Bso</i> CI, <i>Bsp</i> 1286I, <i>Bsp</i> LS2I, <i>Mh</i> II, <i>Nsp</i> II
<i>Sgf</i> I	GCG_AT^CGC	<i>Asi</i> SI
<i>Sph</i> I	G_CATG^C	<i>Bbu</i> I, <i>Pae</i> I, <i>Pfa</i> AIII, <i>Rsp</i> LKI, <i>Spa</i> HI
<i>Sse</i> 8387I	CC_TGCA^GG	<i>Sbf</i> I, <i>Sda</i> I
<i>Tai</i> I	_ACGT^	<i>Tsc</i> I, <i>Tsp</i> 49I
<i>Tau</i> I	G_CSG^C	-
<i>Tsp</i> 4CI	AC_N^GT	<i>Bst</i> 4CI, <i>Hpy</i> CH4III, <i>Taa</i> I

In the preferred embodiment shown in Figure 1, an aminopropanol moiety is at the 3' end of the PRL. This group can be used to covalently attach DNA oligonucleotides to glass surfaces and form arrays [23]. A biotin moiety is bound to the 5' end of the PRL (Figure 1). Biotin and streptavidin have a strong affinity for each other and are routinely used as stable linking agents. Bioluminescent enzymes can be conjugated to streptavidin. When streptavidin-enzyme conjugate

and biotin-labeled PRLs are incubated together, the bioluminescent enzyme becomes attached to the PRLs. Polythymine nucleotide spacers are located at the ends of the PRL (Figure 1). The number of thymine nucleotides in these spacers can be adjusted to decrease steric hindrances of restriction enzyme activity. Other organic moieties, such as alkane chains, could be used as spacers.

This embodiment demonstrates the utility of this invention for determining the presence or absence of specific RNA or DNA sequences (Figure 2). The invention will result in a product that could be supplied as an array of bound, bioluminescently-labeled PRLs. Each PRL would have a stem structure where two complementary regions of the same molecule hybridize. The array will have the streptavidin-bioluminescent enzyme conjugate attached to the PRLs. DNA and/or RNA will be purified by a clinician according to standard methods, which are available in several nucleic acid purification kits. The purified nucleic acids may be sheared into smaller pieces suitable for hybridization and incubated in a buffered solution with the PRL array for hybridization to occur. The temperature and conditions of the hybridization step will be determined by the stability of the possible individual hybrids in the PRL array. A temperature that reduces non-specific binding of target sequences that are not completely complementary to the array loop sequences can be determined for each array of PRLs. After hybridization occurs, the array will be washed and restriction enzyme buffer and restriction enzyme will be added. During incubation with the restriction enzyme solution, the PRLs that are in a hairpin configuration will be cut inside the stem sequence. This cleavage will leave only a few hydrogen bonds holding the label to the surface. The bonds will be unstable and break at room temperature, releasing the label from the bound end of PRL. The array will then be washed again to remove the unbound labels and restriction enzyme. The remaining bound bioluminescent

enzyme labels will be detected by incubating the array with the appropriate substrate solution and measuring photon production. Photon production at a probe site signifies that the sequence within the loop portion of the probe is complementary to a sequence in the target DNA or RNA. Lack of photon production at a PRL sight indicates the absence of a sequence in the target DNA or RNA that is complementary to the probe's loop sequence.

The simplicity and speed of this assay should allow routine use in areas where a parallel genetic test would be advantageous but is often too difficult or time-consuming. Examples of uses include genetic screening for human inherited diseases, determining identity or parentage, identifying pathogens and their antibiotic resistance capabilities, and screening foods or materials for contaminating organisms and viruses.

Other embodiments would include other labels, such as radioisotopic, isotopic, enzymatic, chromogenic, fluorescent, chemiluminescent, or primary and secondary antibodies that are either attached to the PRL during synthesis or added in a later step (before or after hybridization or restriction digestion have occurred). Many new methods for labeling, such as multiple labels, fluorescent beads, and microparticles, are currently being developed and may be used with PRLs in the future [24].

The PRLs can be attached to glass or silicon surfaces according to several other methods [25]. PRLs can also be attached to gold or other metal surfaces via sulfhydryl moieties and to microtiter wells via a biotin-streptavidin linkage [26].

Labels and immobilization moieties can be located anywhere along the length of the PRL, including the loop sequence region, providing that their location does not interfere with hybridization and that the restriction site lies between the surface-attachment site and the position of the label.

The PRL may have different types of spacers or multiple spacers for the purpose of raising the PRL above the surface; to reduce steric hindrances during the hybridization, washing, restriction enzyme digestion; etc. In addition, oligonucleotides can be constructed for use as controls that always remain or are always removed from the surface during the assay.

Literature Cited

1. W. Bains and G.C. Smith. *J. Theor. Biol.* **135**:303-307 (1989).
2. W. Bains, *Genomics* **11**:294-301 (1991).
3. E.M. Southern, U. Maskos, and J.K. Elder. *Genomics* **13**:1008-1017 (1992).
4. U. Maskos and E.M. Southern. *Nucleic Acids Res.* **20**:1679-1684 (1992).
5. S.P.A Fodor, J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu, and D. Solas. *Science* **251**: 767-773 (1991).
6. K.R. Khrapko, Y.P. Lysov, A.A. Khorlin, I.B. Ivanov, G.M. Yershov, S.K. Vasilenko, V.L. Florentiev, and A.D. Mirzabekov. *DNA Sequencing and Mapping* **1**:375-388 (1991).
7. M.D. Eggers, M.E. Hogan, R.K. Reich, J.B. Lamture, K.L. Beattie, M.A. Hollis, D.J. Ehrlich, B.B. Kosicki, J.M. Shumaker, R.S. Varma, B.E. Burke, A. Murphy, and D.D. Rathman. in *Advances in DNA Sequencing Technology*, SPIE **1891**, (1993).
8. C.R. Cantor, A. Mirzabekov, and E. Southern. *Genomics* **13**, 1378-1383 (1992).
9. S. Tyagi and F.R. Kramer, *Nature Biotech.* **14**:303-308 (1996).
10. A. Piatek, S. Tyagi, A. Pol, A. Telenti, L. Miller, F. Kramer, and D. Alland, *Nature Biotechnology* **16**:359-363 (1998).
11. A. Chenn, D. Dorfman, C. Walsh, and J. Winkelman, *Amer. J. of Clinical Pathology* **112**:1 (1999).
12. B. Giesendorf, J. Vet, S. Tyagi, E. Mensink, F. Trijbels, and H. Blom, *Clinical Chemistry* **44**(3): 482-486 (1998).
13. I. Täpp, L. Malmberg, E., Rennel, M. Wik, and A.-C. Syvänen, *Biotechniques* **28**:732-738 (2000).
14. D.L. Sokol, X. Zhang, P. Lu, and A.M. Gewirtz, *Proc. Natl. Acad. Sci. U.S.A.*, **95**:11538-11543 (1998).
15. T. Matsuo, *Biochimica et Biophysica Acta*, **1379**:178-184 (1998).
16. S. Tyagi, "Detecting mRNAs in Living Cells with Molecular Beacons", Contract R01ES10536, National Institutes of Health Environmental Genome Project.
17. X. Fang, X. Liu, S. Schuster, and W. Tan, *J. Am Chem. Soc.* **121**(12):2921-2922 (1999).
18. L.J. Brown, J. Cummins, A. Hamilton, and T. Brown, *J. Chem. Soc., Chem. Commun.* **2000**:621-622 (2000).
19. J.K. Actor, T. Kuffner, C.S. Dezzutti, R.L. Hunter, and J.M. McNicholl, *J. Immunol. Meth.* **211**:65-77 (1998).

20. Venetianer, P. and A. Kiss (1981) In: *Gene Amplification and Analysis, Volume 1: Restriction Endonucleases*, J. Chirikjian, ed. (Elsevier North Holland, Inc.) 209-215.
21. Blakesley, R. (1981) In: *Gene Amplification and Analysis, Volume 1: Restriction Endonucleases*, J. Chirikjian, ed. (Elsevier North Holland, Inc.) 1-34.
22. Sambrook, J., Fritsch, E., and T. Maniatis (1989) *Molecular Cloning: A Laboratory Manual, 2nd Edition* (Cold Spring Harbor Laboratory Press) 5.3-5.32.
23. Maldonado-Rodriguez, R., Espinosa-Lara, M., Loyola-Abitia, P., Beattie, W., and K. Beattie (1999) *Molecular Biotechnology* 11:13-25.
24. Kricka, L. (1999) *Clinical Chemistry* 45: 453-458.
25. Subramanian, A., Kennel, S., Oden, P., Jacobson, K., Woodward, J., and M. Doktycz (1999) *Enzyme and Microbial Technology* 24: 26-34.
26. Fritz, J., Baller, M., Lang, H., Rothuizen, H., Vettiger, P., Meyer, E., Guntherodt, H.-J., Gerber, C., and J. Gimzewski (2000) *Science* 288: 316-318.